Predicting ALS Progression with Bayesian Additive Regression Trees

Lilly Fang and Lester Mackey

November 13, 2012
The ALS Prediction Prize

- **Challenge:** Predict progression of ALS over time
  - Distinguish fast from slow progressors
- **Measure:** ALS Functional Rating Scale (ALSFRS)
  - Score ranges from 0-40
  - Based on 10 questions (Speech, Dressing, Handwriting, ...)
  - Rate of progression = slope of ALSFRS score

- **The Data**
  - 918 training + 279 test patients
    - 12 months of data (demographic, ALSFRS, vital statistics, lab tests)
    - Time series: roughly monthly measurements
  - 625 validation patients
    - Given first 3 months of data
- **Goal:** Predict future ALSFRS slopes for validation patients
  - Error metric: Root mean squared deviation (RMSD)
Outline

- **Featurization**
  - Static Data
  - Temporal Data

- **Modeling and Inference**
  - Bayesian Additive Regression Trees

- **Evaluation**
  - BART Performance
  - Feature Selection
  - Model Comparison
Featurization

- **Goal**: Compact numeric representation of each patient
  - Features will serve as covariates in a regression model
  - Most extracted features will be irrelevant
  - Rely on model selection / methods robust to irrelevant features
**Featurization**

- **Goal:** Compact numeric representation of each patient
  - Features will serve as covariates in a regression model
  - Most extracted features will be irrelevant
  - Rely on model selection / methods robust to irrelevant features

- **Static Data**

  - **Demographics** Age, Race, Sex
  - **ALS History** Time from onset, Site of onset
  - **Family History** Mother, Father, Grandmother, Uncle...

  Categorical variables encoded as binary indicators
Featurization

- **Goal:** Compact numeric representation of each patient
  - Features will serve as covariates in a regression model
  - Most extracted features will be irrelevant
  - Rely on model selection / methods robust to irrelevant features

- **Time Series Data**
  - Repeated measurements of variables over time
    - ALSFRS question scores
    - Alternative ALS measures (forced and slow vital capacity)
    - Vital signs (weight, height, blood pressure, respiratory rate)
    - Lab tests (blood chemistry, hematology, urinalysis)
  - Number and frequency of measurements vary across patients
Featurization

- **Goal**: Compact numeric representation of each patient
  - Features will serve as covariates in a regression model
  - Most extracted features will be irrelevant
  - Rely on model selection / methods robust to irrelevant features

- **Time Series Data**
  - Compute summary statistics from each time series
    - Mean value, standard deviation, slope, last recorded value, maximum value...
  - Compute pairwise slopes (difference quotients between adjacent measurements)
    - Induces a derivative time series
    - Extract same summary statistics
Featurizing Time Series Data

ALSFRS Score

Months
Featurizing Time Series Data

Features extracted
- Mean = 38.75
- SD = 0.816
- Max = 40
- Min = 37
- Last = 37
- etc.
Featurizing Time Series Data

Features extracted
- Mean = 38.75
- SD = 0.816
- Max = 40
- Min = 37
- Last = 37
- Slope = -1
- etc.
Featurizing Time Series Data
Featurizing Time Series Data

Derivative time series

ALSFRS Score

ALSFRS Slope

Months
Featurizing Time Series Data

ALSFRS Score

ALSFRS Slope

Derivative time series

Months

0 0.5 1 1.5 2 2.5 3 3.5

36 37 38 39 40

slope -1

slope 0

slope -2
Featurizing Time Series Data

Features extracted
Mean = -1
SD = 1
Max = 0
Min = -2
Last = -2
Slope = -0.5 etc.

Derivative time series

ALSFRS Score vs Months

ALSFRS Slope vs Months
Featurizing Time Series Data

- **435** temporal features extracted

**Problem: Missing data**
- Average patient missing **10%** of features
- One patient **missing 55%** of features!
- Missing values imputed using median heuristic

**Problem: Outliers**
- **Nonsense values**: Number of liters recorded as MDMD
- Units incorrectly recorded \(\Rightarrow\) **Wrong conversions**
- Extreme values
  - Treated as missing if > 4 standard deviations from mean
Regression model

Future ALSFRS Slope = f(features) + noise

Goal: infer f from data
- Bayesian: Place a prior on f, infer its posterior
- Bonus: Uncertainty estimates for each prediction

What prior?
- Flexible and nonparametric
  - Avoid restrictive assumptions about functional form
- Favor simple, sparse models
  - Avoid overfitting to irrelevant features
Bayesian Additive Regression Trees

- \( f(\text{features}) = \text{sum of “simple” decision trees} \)

- **Simplicity** = tree depends on few features
  - Irrelevant features seldom selected
- Similar to frequentist ensemble methods
  - Boosted decision trees, random forests

*Chipman, George, and McCulloch (2010)
BART Inference

- **Estimating** $f$: Markov Chain Monte Carlo
  - R package ‘bart’ available on CRAN
  - 10,000 posterior samples: $\hat{f}_1, \hat{f}_2, \hat{f}_3, \hat{f}_4, ...$
  
  $$\hat{f}_i = \sum_{i=1}^{100} \sum_{j}^{100} \text{trees}$$

  - **10 minutes** on MacBook Pro (2.5 GHz CPU, 4GB RAM)

- **Prediction**: Posterior mean
  - Average of $\hat{f}_1(\text{features}), \hat{f}_2(\text{features}), \hat{f}_3(\text{features}), ...$

- **Variance reduction**
  - Average predictions of 10 BART models
Accuracy of BART Inference

Number of BART Samples

Validation RMSD

1 sample: 0.5459
100 samples: 0.5234
2000 samples: 0.5144
10000 samples: 0.5109
BART Feature Selection

Top Ten Features Ordered by BART Usage

- Onset Delta
- Max Dressing Score
- ALSFRS Slope
- Last Systolic Blood Pressure Slope
- Mean Weight Slope
- Last FVC Slope
- Last Weight Slope
- Last ALSFRS
- Min Turning Score
- Mean ALSFRS

- Many pairwise slope features
- Lab data excluded
BART on Feature Subsets

Effect of Adding Each Feature in Order of BART Usage

Validation RMSD

Features Added in Order of Usage

- max.dressing
- Onset.Delta
- last.slope.bp.systolic
- alsfrs.score.slope
- mean.slope.weight
- last.slope.fvc.liters
- last.alsfrs.score
- last.speech
- last.handwriting
- meansquares.speech

1 feature: 0.5291
3 features: 0.5246
6 features: 0.5190
14 features: 0.5157
21 features: 0.5113
## Model Comparison

How do other models perform using our feature set?

<table>
<thead>
<tr>
<th>Model</th>
<th>Our RMSD (Test)</th>
<th>Our RMSD (Validation)</th>
<th>Competitor RMSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lasso Regression</td>
<td>0.5006</td>
<td>0.5287</td>
<td>-</td>
</tr>
<tr>
<td>Random Forests</td>
<td>0.5052</td>
<td>0.5120</td>
<td>0.52-0.53</td>
</tr>
<tr>
<td>Boosted Trees</td>
<td>0.4940</td>
<td>0.5118</td>
<td>-</td>
</tr>
<tr>
<td><strong>BART</strong></td>
<td><strong>0.4860</strong></td>
<td><strong>0.5109</strong></td>
<td>-</td>
</tr>
</tbody>
</table>

- **Additive decision tree** models especially effective
- **Featurization** is a main differentiator of competitors
The End

Questions?
Onset Delta vs. Target

Onset.Delta versus ALSFRS Slope on Train and Test Data
Past ALSFRS Slope vs. Target

alsfrs.score.slope versus ALSFRS Slope on Train and Test Data
Last Systolic BP Slope vs. Target

last.slope.bp.systolic versus ALSFRS Slope on Train and Test Data
Max Dressing Score vs. Target

max.dressing versus ALSFRS Slope on Train and Test Data
Mean Weight Slope vs. Target

mean.slope.weight versus ALSFRS Slope on Train and Test Data